OCEANS AND ATMOSPHERE
www.csiro.au

Close-Kin Mark-Recapture for SBT: options for the longer term

Mark Bravington

August, 2014

Consultant Report for the CCSBT

Close-Kin Mark-Recapture for SBT: options for the longer term

Mark Bravington (CSIRO): August 2014

with thanks to Campbell Davies, Pete Grewe, Rich Hillary, and Ann Preece

Close-Kin Mark-Recapture (CKMR) is ready to become an inexpensive tool for directly monitoring future trends in absolute abundance of adult SBT, without needing CPUE or even catch data. Long-term sampling levels can be considerably lower than we needed in the 2006-2010 study: maybe around 1500 SBT total per year, though the appropriate figure will depend on exactly how CCSBT decides to use the CK results for management. If the adult population does increase, then in order to maintain precision it will be necessary to increase the annual sample size too, but more slowly.

Note that, although CKMR does not require catch or CPUE, its successful use for monitoring SBT does rely on continued collection of length- and age-composition data from the Indonesian fishery. And to get proper value out of the CK data in a full stock assessment, a modified OM that directly handles adult length and sex (not just age) will be required.

To safeguard the long-term comparability of the genetic data, now is the time to switch to a more robust and cheaper genetic technology. By moving to "DArT SNPs" instead of microsatellites, we would be able to (i) maintain the same precision with lower sample sizes and total cost, and (ii) identify Half-Sibling Pairs (HSPs) as well as Parent-Offspring Pairs (POPs). With these two types of close-kin, it is no longer necessary to assume that daily catchability is constant with adult size, which we have previously had to do when using POPs alone. It becomes possible to separately estimate the fecundity-at-size and selectivity-at-size curves, and to directly estimate adult mortality. All in all, this would lead to considerably more reliable conclusions about trends in *effective* "spawning stock".

This is a remarkable outcome from a program whose direct costs in the medium term are likely to be well below $100K per year.

1. Introduction

Close-Kin Mark Recapture (CKMR) is a new approach to estimating adult abundance and demographic parameters. It uses modern genetics to identify close relatives (mainly, parents and offspring) amongst large sample sizes of fish, and then makes demographic inferences from the number and pattern of pairs found. The estimates are "fishery-independent", in that they do not need CPUE or total catch data, though they may require length and age compositions and/or supporting biological data. Thus CKMR can be used as an independent check on conventional stock assessments when there are concerns about the validity or interpretation of catch or CPUE

data. Further, assuming the check is passed, CKMR data can be incorporated into a stock assessment in much the same way as tagging data from individual mark-recapture studies.

The first application of CKMR has been to SBT, using genetic samples collected over 2006-2010 from adults in the spawning grounds off Indonesia, and 3yo juveniles in the Great Australian Bight. The method seems to have worked well (Bravington et al., 2014; CCSBT 2012; CCSBT, 2013); it gives much better precision on the adult abundance estimate, rules out certain scenarios previously considered by the OM that are inconsistent with the CK data, and demonstrates that the age-specific pattern of reproductive contribution (ie the appropriate definition of "Spawning Stock Biomass") is quite different to what was previously assumed. The CK data have now also been incorporated into the OM (Hillary et al., 2012) and will form part of future assessments. From 2011-2014, further genetic samples have been collected, but not yet genotyped. The option clearly exists to use CKMR not just as a one-off abundance estimation method, but as an ongoing data stream to monitor adult abundance. This is appealing because the SBT management target is to rebuild the spawning stock, and apart from CKMR there is no other proposed method of directly monitoring that stock.

This paper considers the potential utility of CKMR for future monitoring of SBT adult abundance. It forms the main deliverable of a a research contract under the CCSBT SRP, as proposed at the 2013 ESC. The sections are as follows:

- Genotyping options (now a wider choice than in 2006 when the original study began)
- Utility and practicality of Half-Sibling Pairs, as well as Parent-Offspring Pairs
- A simple model for exploring CKMR design
- Long-term approximate sample size requirements
- Medium-term sampling strategies and possible CVs
- Conclusions and suggestions

2. Genotyping options

Genetics in 2014 uses a bewildering variety of terms to describe different types of genetic marker (marker <=> locus), and to describe different ways of assessing the contents of each locus in each animal (assessing a locus <=> genotyping; several terms already!). There are many terms which, for anybody except a career geneticist, mean almost the same thing. This paper is not aimed at geneticists, so I have erred on the side of simplicity by choosing one word for one group of related concepts, and by trying not to over-explain. Some of my choices might raise the eyebrows of bona fide geneticists.

In 2006 when the CSIRO/FRDC SBT CKMR project began, microsatellites were the only reliable and affordable approach to large-scale "pedigreeless"[1] exercises in finding POPs (Parent-Offspring Pairs) of which that project is still the only example, although not for long). Although other technologies were widely discussed, their reliability and cost was not proven. In 2014, though, there are proven better approaches that use different types of markers (SNP[2]s; see 2.2 for details), and if we were starting the project now, we would not choose microsatellites again. Apart from microsatellites, there are two quite different SNP approaches— SNP-chips, and DArTs— and we consider all three approaches separately. Note that the previous study only consider POPs, but this paper considers Half-Sibling Pairs (HSPs) too, and the term "Close-Kin" here covers both types of relative.

[1] IE no prior information about parents.
[2] SNP = Single Nucleotide Polymorphism, if it helps to know that.

It seems clear (and is demonstrated later in this document) that CKMR is going to be an important part of SBT management in future, and this begs the question of whether and when to change to a better genetic technology. There are three factors to consider:

- reliability and "future-proofing"
- cost
- information content

A fundamental requirement for any approach, is the ability to *reliably find CK pairs*, i.e. to exclude almost all false positives and to generate hardly any false negatives[3]. The demonstration that we had achieved this with 25 microsatellites, was a laborious but crucial part of the original SBT POP analysis. However, there is a simple key message, regardless of which approach: use *enough* loci, and build in a few extra to be on the safe side. Since SNPs are individually much less informative about likely relatedness than usats, clearly many more SNPs are required. In this paper, I have simply assumed that enough loci *will* be used, as appropriate to whichever genetic approach is taken. For further discussion of exactly how many loci is enough, see Appendix D.

2.1. Microsatellites

[for brevity, written as "usats" and pronounced as "microsats"]

Microsatellites have worked well so far for SBT, but they are not a good bet for the long-term. The act of genotyping with usats involves considerable subjective skill on the part of the "reader" (even with up-to-date automated software), and readers need to be trained up for each new species and painstakingly intercalibrated. Also, if some of the chemical steps are outsourced (as we did for SBT, to save time and money), then the staff at outsourced end need to be well-acquainted with the protocols and the nuances of each species— a lesson learnt from experience. Further, results depend on the actual technology platforms used, and these are not guaranteed to exist for ever.

In general, usats have become much less popular in recent years. The following quote from Fernández et al. (2013)— which refers to "STR[4]s" instead of usats— explains why:

> "Despite being highly polymorphic, informative and interspersed throughout the entire genome (Baumung et al., 2004; Tian et al., 2007), the results obtained with STRs by different laboratories are not always comparable because of inconsistencies in allele size calling and errors in size determination. Furthermore, STRs are time consuming for trained personnel to analyze, even with the use of appropriate software or other automated methods for allele analysis (Vignal et al., 2002). Recent advances in high-throughput DNA sequencing, computer software and bioinformatics have made the use of SNPs more popular (Heaton et al., 2002). Although in terms of genetic information a biallelic marker may be considered as a step backwards, SNPs have some promising advantages, including greater abundance (Heaton et al., 2005), genetic stability in mammals (Markovtsova et al., 2000; Nielsen, 2000; Thomson et al., 2000), simpler nomenclature and suitability to automated analysis and data interpretation (Wang et al., 1998; Lindblad-Toh et al., 2000)."

Here is another relevant quote (Hauser et al., 2011):

[3] False-positives are unrelated pairs that get accidentally classified as close-kin, because they happen by chance to be genetically similar enough to pass whatever classification criterion is used. False-negatives are true close-kin that by chance fail the criterion.

[4] Short Tandem Repeats, if it helps to know that.

> "On the other hand, SNPs provide the advantage of transferability between laboratories, suitability for database entry (Hauser & Seeb 2008), high- throughput and low genotyping error (Seeb et al. 2009). For example, collection of the microsatellite data for the present study took 3 months, while SNP data were collected in a week."

These and other similar comments (e.g. Seeb et al., 2011) date from several years ago; since then, SNP technology has become even more attractive and cheaper by comparison.

2.2. SNP-chips

[The preferred formal term is "targetted SNP assays"]

For CK purposes, a SNP is a locus with only two possible alleles A and B, in contrast to usats which (if suitable for CK) will jave ten or more. Each animal at each SNP locus is either AA, AB, or BB. Close-kin pairs are more likely to have one or two alleles in common. However, because of the limited range of possibilities, individual SNPs are not very informative about kinship, so many more SNPs than usats are needed for CK. Fortunately, SNPs are individually far cheaper to genotype.

With the SNP-chip approach, a set of SNP loci are specially chosen for inclusion on a mass-produced "chip", with each chip handling about either 50 or 100 SNPs for a batch of around 100 animals. For individual ID, e.g. in gene-tagging, a single chip with about 50 SNPs is quite adequate, but for reliably finding POPs, at least 200 SNPs— i.e. 2 or 4, maybe 5 chips— would be needed (see Appendix D). SNP-chips have been used for some years in the world's by-far-largest wildlife parentage program (Anderson, 2010), to assess the proportion of hatchery-reared salmon along the western USA coast; it is important to realize that the parentage task is much easier in that program than for SBT, since the possible-parents-of-interest (i.e. the hatchery broodstock) have known genotypes. Hence the number of SNPs required in the salmon program is much less than for SBT, where the potential parents are unknown.

There are a number of different proprietory companies making SNP-chips, with considerably varying requirements in infrastructure and running cost; whether all the companies will still be in business in 2025 is another question. Using the same SNPs but in chips from different companies is possible in principle, but would require some effort. On the plus side, there is no concern about subjectivity or repeatability when SNP-chips rather than microsatellites are used; genotyping is reliable and automated. If the number of SNPs could be kept to 200, then unit running cost per fish of SNP-chips for POP purposes would be lower than for equivalent usats, provided the company has been chosen well.

Developing enough SNP loci for a SNP chip used to be a very challenging exercise, akin to developing special-purpose usats. To keep the total SNPs to 200, it is necessary to choose SNPs with the minor allele frequency fairly close to 0.5. Nowadays the burden is somewhat reduced, thanks to the DArT-type approaches discussed next.

2.3. SNPs by DArT

[DArT ("Diversity Array Technology": see `http://www.diversityarrays.com`) is a proprietory term for one of a class of methods sometimes called Next Generation Sequencing, or NGS. DArT gives fewer SNPs per fish than other NGS methods, but each SNP is genotyped more reliably, which is important for CKMR.]

DArT offers thousands-to-tens-of-thousands of SNPs per animal, with almost zero development cost. Per-genotype running costs are currently somewhat higher than for usats and SNP-chips (maybe twice the cost),

but the costs of the sequencing technologies that underlie this approach have fallen massively in recent years and continue to do so; cost reductions in microsatellites and targetted assays have been much lower. DArT and related approaches are heavily outsourced, especially compared to usats which can largely be done in-house. If a good company is chosen, genotyping is reliable and automated. At least in principle, the technology is robust to being transferred to a different provider.

The number of SNPs from DArT vastly exceeds what is needed for POP-finding, but opens the possibility of finding HSPs too, as discussed in section 3. Of the three viable approaches to CK for SBT, DArT is the only one that can identify HSPs.

Over the last two years, CSIRO has been trying DArT on several species of fish, including several tunas. Based on the theoretical power and the practical success, we have decided to use DArT on all new CKMR projects, excluding perhaps the special case of individual ID when animal sample sizes are very large (10,000s), in which case a single SNP chip may be better. So far, our experience with DArT has been very positive—perhaps because we have found a particularly good and co-operative service provider.

DArT SNPs are reputedly somewhat more demanding of high-quality DNA samples than usats or SNP-chips, but in our prelimary run on 90 SBT (all successfully scored with usats), it seems that only one or two fish had inadequate DNA for DArT. We therefore do not expect any significant DNA-quality issue for SBT DArTs.

2.4. Summary

The utility of CK data entirely depends on genotyping being consistent over time. With usats this is always going to be at risk. Reliability of usats is a concern, and long-term costs will be lower moving to SNPs. Although DArT SNPs have a (currently) higher unit cost than SNP chips or usats, DArT provide much higher information content per fish because (as explained next) it can reveal HSPs as well as POP, so fewer fish need be genotyped to achieve a given precision; and, unlike the other methods, DArTs are likely to become much cheaper in future.

Note that the economics look rather different for individual gene-tagging, where unit cost is paramount because sample sizes are very large, but only a small number of SNPs are needed; there, a single SNP chip appears to be the best option.

3. Half-Sibling Pairs (HSPs)

3.1. Why HSPs?

If we used DArTs, we could find HSPs as well as POPs. Setting aside cost and logistic issues for now, there are two main reasons why HSPs as well as POPs are desirable:

- HSPs provide more CKPs per sample, thus addressing one of the main limitations on CV of abundance estimates.

In fact, HSPs will probably triple the total number of CKPs per sample, allowing us to substantially cut the long-term sample size requirement for monitoring.

- With HSPs as well as POPs, we become able to separately estimate (adult) selectivity, fecundity, and mortality.

Until now, with POPs alone, we have had to *assume* a fixed relationship between selectivity-at-length and fecundity-at-length, as determined by external data on daily fecundity and assuming that daily catchability is independent of size. This assumption has substantial implications e.g. for assessing trend in "effective SSB", so it is highly desirable to be able to avoid relying on it. (And of course the assumption is fundamentally incompatible with changes in adult selectivity over time.)

A third, and less important, benefit of HSPs, is that the presence of *large* numbers of *within-year* full-sibs[5] and half-sibs would[6] lead to overdispersion in the POP-based CK data. The evidence to date suggests not (e.g. so far no adult has been involved in more than one POP), but the only direct way to check is to actually look for half-sibs.

The rest of this subsection gives a informal perspective on the utility and potential pitfalls of using HSPs. The rest of section 3 takes a more formal approach, and also addresses the technical viability of finding HSPs; most of the material is deferred to Appendices.

3.1.1. The cartoon version of CKMR

The cartoon version (i.e. heavily simplified, mainly by ignoring time altogether) of CK MR with POPs goes like this. We start by genotyping one specific Juvenile. It must have two parents. If a random Adult is picked from the adult population of size N, then the chance of it being one of the parents of the Juvenile is $2/N$. If we repeat this exercise for each pairwise comparison between m_J juveniles and m_A adults, then the *expected* number of POPs found is $2m_Jm_A/N$, so N can be estimated from $2m_Jm_A/P$ where P is the *actual* number of POPs found.

The analogous cartoon version with HSPs goes like this. We start by genotyping two Juveniles— John and Jane. They each have one mother. The chance that they have the *same* mother is 1 over the number of *female* adults, i.e. $2/N$. Similarly, the chance that they have the same father is $2/N$, so overall the chance that John and Jane are an HSP is $4/N$. Now, if we compare all m_J samples with each other, giving $m_J^2/2$ "non-double-counted" comparisons, and find H HSPs in the process, then we can estimate N from $2m_J^2/N$.

When sampling is split 1:1 between juveniles and adults[7], the cartoon suggests that the number of POPs and HSPs should be about equal, but for many reasons this need not apply in practice. Most notably, after an adult has been caught, it can no longer produce offspring, so the POP data-series is "one-way" in time; in contrast, catching a Juvenile does not stop its parent from producing more half-siblings of that Juvenile. The simulations for SBT shown later in this document indeed suggest that about twice as many HSPs as POPs would be found under 1:1 sampling.

3.1.2. Variation in fecundity/reproductive success

A FAQ about POP-based CK MR is: but is it affected by variation in reproductive success? The short answer, at least in the cartoon world, is "no" (Figure 3.1); there is no bias, though variance would be affected by high RV since it would no longer be true that the comparisons are (almost) statistically independent. Having said that, in the non-cartoon world of SBT, fecundity (which varies systematically through an adult's life, as it grows)

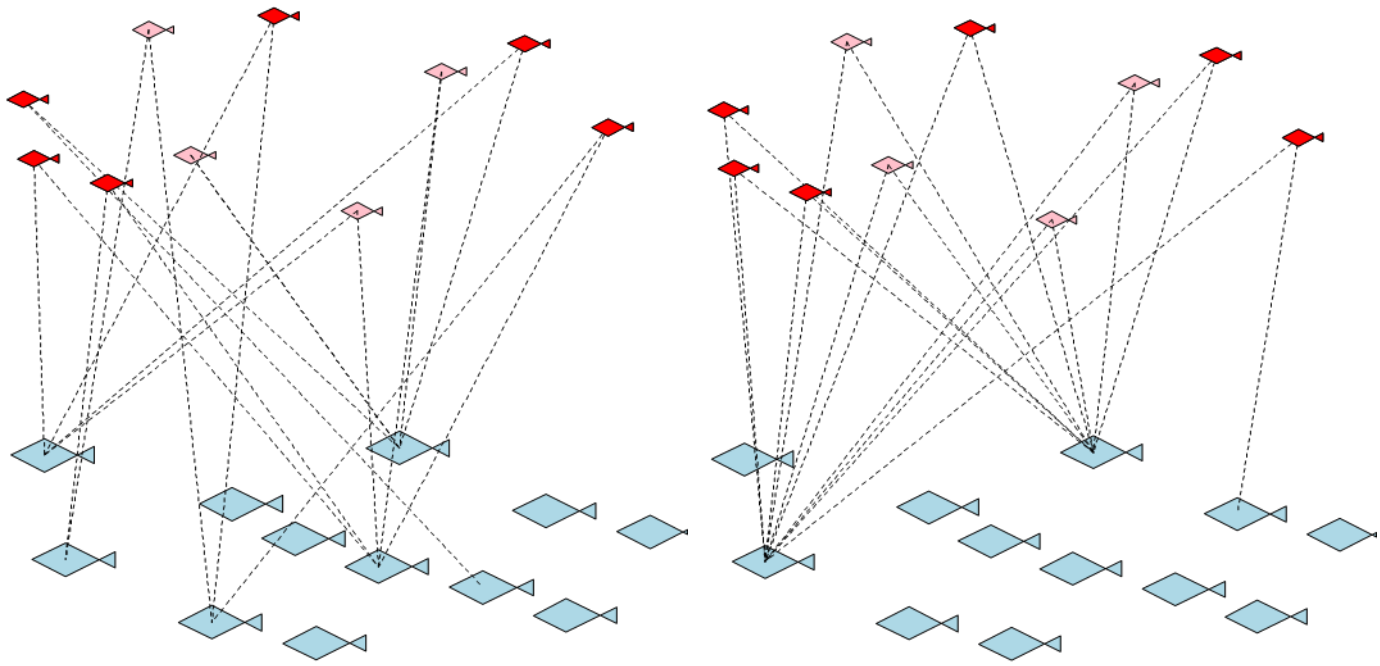
[5] Identifying full-sibs is much easier genetically than identifying half-sibs, so if we can do the latter then we can certainly do the former.

[6] It is inevitable that there will be *small* numbers of HSPs within each cohort of juvenile samples, but only of the same order of magnitude as the proportion in POPs, which is very low.

[7] In this paper, a ratio M:N always means "future ratio of genotyped juveniles to adults"]

is important in analysing CK MR data, and can be allowed for; but year-to-year fluctuations in reproductive output make no difference.

Figure 3.1: Non-impact of reproductive variability on POP-based CK MR



Low RV on the left; high RV on the right. However, the number of links in both pictures is the same. Each pairwise comparison has an equal chance of finding a link, and the estimate of adult abundance is based on the number of *links* found, not the number of *parents* found.

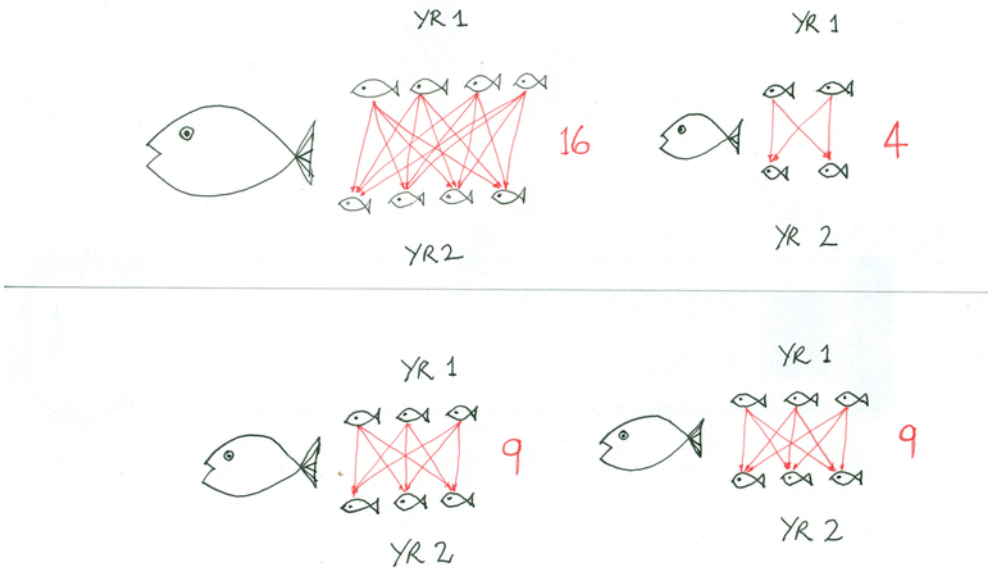
With HSPs, though, variations in reproductive output is much more important. Figure 3.2 shows how systematic variation in fecundity (i.e. *average* reproductive output per year) affects the number of HSPs present. There are 20 cross-cohort HSPs (ie only comparing juveniles in *different* cohorts) in the top scenario, but only 18 in the lower. The same is true within-year; in the Figure, there are 30 comparisons per row (within-cohort) of juveniles, yielding 21 within-cohort HSPs in the upper scenario, but only 16 in the lower.

It is certainly possible that random within-year fluctuations in individual reproductive success could affect the number of HSPs found within each sampled cohort of juvenile SBT, so to be on the safe side we exclude within-cohort comparisons when using HSPs. Even in the cartoon world, that still leaves growth-related fecundity to be dealt with. There is no way to allow for fecundity using HSPs on their own (because HSPs give no information on which adults were responsible for them), but of course for SBT we have the POPs as well, from which the age- (or size-)specific fecundity can be estimated. Roughly speaking, this allows "independent" estimates of abundance to be made from the number of HSPs, as well as from the number of POPs. More accurately, the two abundance estimates are only independent once the structural parameters (fecundity, mortality, selectivity) are estimated, as described in the next subsection. The combination of the two types of CKP is more powerful than either alone.

3.2. Independent estimation of selectivity, fecundity, and mortality via HSPs and POPs

The interpretation of POPs (ie the number found, and their patterns with age and time) is affected not just by adult abundance, fecundity, and mortality, but also by the complication of adult selectivity. To get round this for

Figure 3.2: Impact of reproductive variability on HSP-based CKMR



Variable fecundity scenario on top; constant fecundity below. Bigger fish make more babies. The upper and lower sets of juveniles *within* each scenario represent different cohorts.

SBT CKMR, it has until now been necessary to assume that selectivity (within sex) is directly proportional to residence time on the spawning grounds, ie that catchability-per-day is independent of length (within sex). By making this assumption, and combining it with external data on daily spawning output as a function of (female) size, we enforce a "hard link" between selectivity and fecundity which allows them both to be estimated.

While it is hard to argue with the notion that residence time on the spawning grounds must be a primary driver of selectivity for SBT caught on the spawning grounds, it is also hard to argue that it must be the *only* important driver. The problem is that there has been no data to help; it *might* eventually be possible to e-tag enough adult SBT (recaptured later through the Indonesian fishery) to determine residence time directly; this has recently been done for Atlantic Bluefin Tuna by Aranda et al. (2013), although unfortunately they could not accurately record body size. However, for SBT such a tagging program would take a long time and be expensive.

The basic ingredients for estimating structural (non-abundance) parameters from SBT POPs are age compositions (in adults, and in identified parents), and the mean time-delay between catching an offspring and its parent. It turns out that adult selectivity and mortality are inextricably linked in all these. In contrast, selectivity obviously does not affect HSPs at all. We therefore have some grounds to hope that HSPs and POPs together might be able to disentangle the parameters, without requiring the equal-daily-catchability assumption. The next paragraph gives a "heuristic" explanation of why this is possible.

Consider what happens after we catch one juvenile. As time goes by afterwards, how does the probability change that any one comparison reveals (i) a Parent, and (ii) a Half-Sibling? In both cases, the probability diminishes at the mortality rate because the parent might die, and also diminishes at the the population rate-of-increase because there are more non-Parents and more non-Half-Siblings around. Assuming that the parent does survive and is still around to be captured and/or make babies:

- For POPs: the parent becomes more catchable each year because of increasing selectivity with age
- For HSPs: more half-sibs are available for capture each year, because the parent becomes more *fecund*

over time.

These trends in probability determine the mean time gap for POPs and for HSPs. The difference between the mean gaps for the two types of CKP thus reveals the difference between age-specific selectivity and fecundity. Once these two are distinguished, it becomes possible to estimate all the remaining structural parameters including adult mortality, based on age compositions and the trend over time in CKPs-per-year.

More formally, Appendix A demonstrates that, in a deliberately simple and abstract setting, all parameters do become estimable if and only if HSPs and POPs are used together. In the Appendix A, selectivity and fecundity are assumed to be linear functions of log(age), and there is no natural contrast from year-to-year recruitment fluctuations. These assumptions are of course not meant to be realistic— they are to illustrate a point. In practice, when the dependence on age is nonlinear it does seem to be technically just possible to separately estimate fecundity and selectivity from POP data alone, in the sense that the Fisher information matrix is invertible and finite (but rather large) CVs are obtained (section 6). Presumably, though, this relies on subtle features of the assumed functional forms, and may be sensitive to model choice.

Note that, although the HSPs themselves only involve juvenile fish, HSP-based CKMR still provides information *only about adults*. There is no way to to use CKMR to estimate anything useful about juveniles[8].

3.3. Can HSPs be reliably distinguished with DArTs?

Yes for SBT — at least, according to statistical theory and our preliminary SBT DArT data, though we cannot prove it yet. As of 3/8/2014, we have DArT results from about 95 SBT, including almost all the 90 POP-members identified in the original CK study. Given the number and quality of SNPs found, the preliminary statistical analysis in Appendix D shows that almost all HSPs should be reliably identifiable. However, we do not yet have any identified HSPs— and we never will have any from existing usats, because the 25 usats that we use for POPs are not enough to identify HSPs. Therefore it will be important to confirm that we really can find HSPs in SBT before switching completely to DARTs. The only real way to do so, is to genotype enough juveniles to generate enough juveniles to be demographically sure that a respectable number of HSPs is present. Appendix D has more to say on the actual process; on demographic grounds, 3000 SBT juveniles should contain on average 25 or more HSPs (here we can use within-cohort HSPs as well as between-cohort HSPs). The information gleaned from genotyping the 3000— basically, the variance of the measured relatedness across HSPs— would then allow us to refine and cost-prune the DArT process for subsequent SBT, by trimming superfluous loci.

After an HSP is identified, it is also very useful to know whether the shared parent is the mother or the father. This can be done, with at least some degree of statistical precision, by (separately) examining the mitochondrial DNA of all HSP members (i.e. a small percentage of all genotyped juveniles).

4. The model for this CKMR design

To do the calculations for this paper, I found it necessary to implement a new "CK assessment model" (Appendix E); this is partly to cope with HSPs, but also to speed up calculation and to circumvent some structural assumptions that may be undesirable. The original CK Mini-Assessment (CK-MA; Bravington et al., 2014),

[8] At any rate, in the 8 years that I've been thinking about CKMR, I have been unable to come up with any way to use CKMR to estimate juvenile abundance or survival, except for one limited application in an enclosed river system.

which is fully length- as well as age-structured, is too complicated and slow for design purposes, and also gave baseline estimates with a much stronger adult abundance trend[9] than the CK-OM combination model (Hillary et al., 2012) suggested. Since the recent trend has a big impact on future projections, it seemed wiser to start the projections with recent estimates from the CK-OM which are presumably thought to be better since they use other information. Some of the structural assumptions of the CK-MA may also be inadequate for long-term work, e.g. constant selectivity over time. Unfortunately, the CK-OM itself is not suitable for CK projections either; the most obvious problem is that it lumps both sexes together (in an admittedly ingenious way), despite plentiful sex data, clear evidence of dimorphic adult growth, and every reason to expect different selectivity and fecundity patterns by sex!

Instead, for this paper I developed a new Approximate Mini-Assessment (AMA) for adult population dynamics only, using just CK data and adult age data. It is purely age-and-sex-based model, with constant mortality z over time and over ages 8-24, a plus-group at[10] 25 with its own z_+, fixed sex-ratio over time, and sex-specific selectivity and fecundity curves. The model was seeded with a hybrid of estimates from the CK-MA and the CK-OM. Details are given in Appendix E. CVs were computed from the CK data *alone* via expected Fisher information (i.e. the right way, based on a proper likelihood), but the (highly informative) Length-and-Age (L&A) data are not used explicitly in forming the likelihood; instead, they are assumed to contain exactly enough information to accurately set the age distribution in the adult catch.

As discussed in Appendix E, this treatment of L&A data is not fully satisfactory. Using just the "old" existing data up to 2010, the new AMA actually gives *higher* CVs on abundance than the old CK-MA did, perhaps not extracting enough information from the L&A; however, in long-term (20+ years) simulations, the new AMA gives CVs on abundance that are implausibly *low*, because the L&A data are being "taken too literally" in fixing relative proportions across many years. For any given time period, the relative CVs from different sampling strategies ought still to be reasonable, and over the medium-term (say 5-10 years) I think the absolute CVs are probably not unreasonable too.

5. Long-term sampling targets

CKMR projects need to reach around 50 CKPs before anything at all precise can be said. It makes sense to try to get to the 50 as quickly as possible, which means heavy sampling in the first few years. However, once this is done and a respectable "back-catalogue" of samples has become available, intuition suggests that annual sample sizes can be reduced in order to use CKMR as a monitoring tool rather than for a one-off abundance estimate. The rationale is that each new sample (i) can be compared against all pre-existing ones, and (ii) then becomes a member of the back-catalogue; this generates a quadratic increase in the rate of finding CKPs. Eventually the quadratic magic stops working, because an adult caught a very long time after a juvenile is very unlikely to be a *surviving* parent of the juvenile; a similar argument applies to half-siblings separated by many years.

To analyse CKMR data for SBT, we need to include not just abundance parameters but also "structural" parameters (selectivity, fecundity[11], and mortality). The structural parameter estimates are somewhat correlated with the abundance parameter estimates, so uncertainty about the structural parameters can can contribute substantially to uncertainty about abundance (and, particularly, absolute abundance) in the short term. However, if

[9] The CK mini-assessment had considerable uncertainty about the *trend* in abundance over 2003-2010, though less so about the mean.

[10] To match the original choice in the CKMA, so that no extrapolation of previous results was required. With hindsight, 30 would have been a better choice in the original CKMA.

[11] The $(z/2)\sqrt{N}$ result only relies on POPs, and unless HSPs are used then selectivity will continue to be inferred "by assumption" relative to fecundity, as in the existing CK mini-assessment.

the project goes on long enough, the structural parameter uncertainty will eventually become negligible as data accumulates, since these parameters are estimated using *all* the data collected. The intrinsic uncertainty about the time-series of abundance will remain, though, because each cohort of adults influences only a limited part of the entire data series and eventually no more relevant data can be collected on that cohort.

A simple theoretical argument (Appendix C) predicts that, in a *heavily simplified* steady-state setting with adult abundance N, the *annual* number genotyped in order to get a CV of C on abundance[12] of C is only about

$$\frac{1}{C}\frac{z}{2}\sqrt{N}$$

per year. For SBT, we have already genotyped about $10\sqrt{N}$ SBT in 4 years, ie about $2.5\sqrt{N}$ per year, whereas z is about 0.1-0.2. So, to the extent that the simplified theory is appropriate to SBT, a 10% CV could could ultimately be achieved with only $0.8\sqrt{N}$ samples per year, i.e. somewhere around 1000 (combined juveniles and adults) per year. While this ball-park figure is useful, there are many reasons why it should **not** be taken too seriously— and, as shown by simulation results below, it is appreciably wide of the mark for SBT.

Long-term means that monitoring has continued long enough to precisely estimate the structural parameters. That will take some years.

Trends in abundance are ignored. By the time we reach "the long term" for SBT, adult abundance will presumably have gone up substantially. That will mean more samples are needed to maintain precision. The **good news**, though, is that the increase in sample size need only be proportional to the square-root of the increase in abundance. Another way to think about that is: if abundance is increasing at $R\%$ per year, then sample size needs to increase at only $0.5R\%$ to maintain monitoring precision. As stock size increases, CK MR becomes cheaper relative to fishery revenue (at constant f), albeit more expensive in absolute terms.

Selectivity/fecundity are ignored, whereas they will in practice affect sampling efficiency even if the parameters are known. If there is an imbalance between selectivity-at-age and fecundity-at-age, and adults are sampled in proportion to their abundance in the catch, then the per-sample rate of POP-finding will be (somewhat) lower than assumed. This could be adjusted for SBT, by preferential genotyping of larger adults, but at the cost of needing to wait longer to accurately estimate recent cohorts. (Note that projections from the AMA do account for selectivity and fecundity.)

Retrospective abundance estimates. The theory is based on the accumulated total offspring found for each cohort of adults, so to achieve full "convergence" (in the sense of a VPA, say) on the abundance estimate for year Y, it is in principle necessary to wait until the youngest adult cohort in Y has delivered its most of its reproductive output— which could be 15 years for SBT. That may be a little too long to wait for management. CVs for the more recent past will be somewhat higher.

Having said that, the theory ignores information on relative recruitment that comes from the L&A data— ages are implicitly used only to assign adults to cohorts for POP-checking. In practice, this information would reduce the CVs on abundance in the recent past.

Target CV is quite arbitrary: why 10%? It depends entirely on the management context— which is far beyond the scope of this report.

[12] Absolute abundance of adults, in any single chosen year substantially before the "present" (ie long before the most-recently-sampled year in the long-term study).

POP-only CK is assumed. Clearly, HSPs would substantially *improve* the precision— the long-term simulations in the next section suggest that equivalent CVs could be achieved with just 60% of the sampling levels if HSPs are used.

The AMA can be used to refine the ball-park figure, at least in terms of fecundity and selectivity effects. Unfortunately, the AMA itself is too simplified to yield meaningful CVs on abundance over very long periods, because it assumes that L&A data is good enough to accurately set the relative abundance of *any* two cohorts regardless of how far apart in time they are (so the abundance estimates from different cohorts never become "decoupled", and the resulting CVs are implausibly low in very long studies). In practice, the L&A data do give useful information on relative abundances of *nearby* cohorts, but will be less informative over the long term because of accumulated errors. Nevertheless, the AMA should still give reasonable predictions of the *number* of POPs and HSPs likely to be found for each adult cohort or year-of-sampling, taking appropriate account[13] of selectivity and fecundity.

With a total of 1000 genotyped animals per year split 2:1 juveniles:adults (i.e. 666:333), and steady-state abundance, the AMA predicts that eventually around 6.9 Parents would be identified per year. This also implies that 6.9 Parents would eventually be identified from each adult cohort (with 6.3 coming from ages 8-24, and the remainder from older animals), and the fully-retrospective CV of initial cohort size for each cohort will be $1/\sqrt{6.9} = 0.39$. To infer adult abundance in the (fairly distant) past, eqn C.2 can be applied with (say) $z = 0.15$, giving a fully-retrospective CV from POPs alone of 0.21. This is almost twice what the "ballpark theory" would suggest, so fecundity and selectivity are clearly important here. Note that the 2:1 sampling split is not optimal for POP-only CKMR (it is close to optimal when using both HSPs and POPs, which is why I used it here), but changing the split to 1:1 has only a modest effect on long-term CV.

It is not immediately obvious how to incorporate HSPs into this long-term approach, but one possible way is simply to note that the expected number of HSPs found in any single year[14] is inverse to adult abundance in that year, with known constant of proportionality (depending on the structural parameters). Since HSPs and POPs are statistically independent, the HSP count in any given year[15] will ultimately provide an independent estimate of adult abundance that year, with a CV of $1/\sqrt{\mathbb{E}[\text{HSP}]}$. With the 666:333 sampling of the previous paragraph, the AMA predicts an annual HSP count of 17. An inverse-variance-weighted combination of the HSP- and POP-based retrospective abundance estimates would (from eqn C.3) have a CV of around 0.16, substantially better than 0.21 in the previous paragraph. This is an unsophisticated way to use the HSPs, and a more nuanced analysis in practice might yield a lower CV. However, it is still considerably more than the evidently-too-simple ballpark theory predicts. To achieve the same long-term fully-retrospective "target" annual CV of 0.10 (or any other target CV), we can simply scale the 666:333 sampling by the ratio of target CV ratio, which translates to 1000:500 sampling if the target is 0.10. (As noted in the CK study, with CKMR the annual sample size is directly inverse to CV, because the number of comparisons scales quadratically).

Even in a fully-retrospective context— which, as noted above, may not be that useful— it turns out that it is HSPs rather than POPs which actually contribute most of the information on adult abundance, despite the POP-based estimates being able to "borrow strength" across cohorts. This is not surprising, given that each year's 1000:500 sampling in the long-term AMA would yield about 38 HSPs and only 15 POPs; this is affected by having split the sampling in favour of juveniles, but there are still about twice as many HSPs as POPs even with 500:500 sampling. In fact, even more CKPs per genotype could be generated by a more extreme split in

[13] Assuming the constant-daily-catchability model is correct...

[14] IE with one member caught in that year, and the other in a different year

[15] IE where the first member is caught in that year, and the second in any subsequent year (to avoid double-counting the past).

favour of juveniles. The downside, though, is that information on parental age only comes from the POPs, and this information is what determines the fecundity-age relationship, and the parameters of that relationship are essential for making an abundance estimate from HSPs, and for actually getting to the "long term" situation addressed in this section. Of course, we do already have 45 POPs, which provides a very useful head-start on fecundity estimation.

6. Medium-term projections

This section examines likely CVs that could be obtained at two medium-term review points (2018 and 2025, ie using data collected up to those years) under various scenarios of CK genotyping, both with and without HSPs. All calculations were done with the AMA described in Appendix E.

The **aspects considered** in Table 1 cover the following (but not all permutations, since the different aspects act largely independently on CV):

- whether HSPs are used or not
  - if yes, whether current juveniles and/or adults are re-genotyped with DArTs
  - if not, whether the constant-daily-catchability assumption is applied in order to hard-link selectivity and fecundity
- different annual sample sizes in future, for juveniles and adults

The **results shown** are CVs for N_{10+}, i.e. total numbers of adults aged 10 and up, in various years, plus the SE of the log-change in abundance between 2006 and 2016. (If the abundance were to double over this period, the log-change would be about 0.70— though the CVs would be worse because there would be fewer CKPs.) The general caveats of Appendix E apply: comparative CVs across sampling strategies are probably realistic, but absolute CVs may be too high or too low in general; CVs for the most recent abundances are probably too precise.

Table 1: CVs (per cent)

REVIEW 2025

	HSP	σ/ϕ link	Regeno J / A	m_J m_A	CV 2010	CV 2017	CV 2024	SEtrnd '06-'16	#POP #HSP	Notes
#1	Y	N	Y	1000	09	12	15	09	119	J+A regeno
			Y	500					456	
#2	Y	N	**N**	1000	11	14	15	10	103	no regeno
			N	500					327	
#3	Y	N	**Y**	1000	09	12	15	09	125	J regeno
			N	500					466	opt J:A if HSP
#4	Y	N	Y	**500**	14	19	25	15	71	half of
			N	**250**					209	above
#5	Y	N	Y	**750**	09	13	16	10	139	opt J:A if
			N	**750**					322	POP-only
#6	**N**	**Y**	**N**	750	11	17	25	17	139	no HSP
			N	750					0	*assumed* σ/ϕ
#7	N	**N**	N	750	20	26	26	17	139	no HSP
			N	750					0	*estimated* σ/ϕ
#8	N	**Y**	N	**1200**	08	11	16	13	259	CV $\approx$ #3
			N	**1200**					0	

REVIEW 2018

	HSP	σ/ϕ link	Regeno J / A	m_J/ m_A	CV 2010	CV 2017	CV 2024	SEtrnd '06-'16	#POP/ #HSP	Notes
#1	Y	N	Y	1000	12	23	-	20	89	J+A regeno
			Y	500					282	
#2	Y	N	**N**	1000	16	26	-	21	70	no regeno
			N	500					168	
#3	Y	N	**Y**	1000	13	24	-	21	83	J regeno
			N	500					282	opt J:A if HSP
#4	Y	N	Y	**500**	18	37	-	32	58	half of
			N	**250**					157	above
#5	Y	N	Y	**750**	13	26	-	22	98	opt J:A if
			N	**750**					214	POP-only
#6	**N**	**Y**	**N**	750	16	37	-	35	98	no HSP
			N	750					0	*assumed* σ/ϕ
#7	N	**N**	N	750	24	41	-	35	98	no HSP
			N	750					0	*estimated* σ/ϕ
#8	N	**Y**	N	**1300**	11	25	-	26	172	CV $\approx$ #3
			N	**1300**					0	

- **Bold** shows what has changed from previous line.
- Leftmost column is just the "scenario number", for reference.
- "σ/ϕ link": is daily catchability assumed to be independent of size? If not, σel.- and ϕec.-at-age are separately estimated.
- "Regeno": re-genotype 2006-2010 Juveniles and/or Adults with DArTs? If not, cannot compare across the 2010 boundary.
- CVs are for *retrospective* abundance estimates to that year, using data up to the Review year.
- "SEtrnd" is 100*standard error of the log-trend (e.g. a true doubling would be 70=100log$_e$ 2 on this scale).
- Assumes no trend in 8yo recruitment over study period.
- Mortality fixed for plus-group at $z_{25+} = 0.29$, matching CK-OM, but estimated for younger adults.
- #8 is a POP-only scenario with sample sizes chosen to give similar CV $\left[\hat{N}\right]$ to scenarios #3 or #5, which use POP+HSP. Required sample size depends on the Review date.

A number of conclusions can be drawn from Table 1.

1. DArTs (ie HSPs) look good. The long-term annual sample size required to get equivalent CVs from POPs alone (#3 vs #8 in each table) is nearly double (1200 vs 750), or more in the shorter term (1300 vs 750), and even then the trend estimate is considerably less precise (26% vs 20% at 2018).

    a) Plus, of course, the with-HSP results do not rely on the constant-daily-catchability assumption. The POP-only results, which (apart from #7) do rely on that assumption, therefore understate the true uncertainty.

2. It is *not* worth re-genotyping adults that have usats already (#1 vs #3). This makes intuitive sense: adults cannot contribute any further POPs more than 3 years after being caught, so the 2006 and 2007 adults are already "used up". The 2008/9/10 adults have already delivered most of the POPs that they ever will, and re-genotyping them will give less bang-for-the-buck than genotyping post-2010 adults instead.

3. It probably *is* worth re-genotyping juveniles that have usats already (#2 vs #3). There is some difference in CV, albeit not dramatic, but more importantly there is a substantial difference in the number-of-POPs and number-of-HSPs obtained by 2018, and of course also in the length of the time series that can be spanned by any HSP or POP (important for estimating structural parameters such as selectivity). This is likely to lead to qualitatively better models in a way that may not be fully captured by these CVs, which are based on a simple model without error checks.

    a) I did not consider the possibility of genotyping more "fresh" juveniles from 2006-2010. This would be another way to quickly increase the number of POPs and HSPs, but we are close to running out of suitably-sized fish (ie that we are confident of being 3yo) in the archives.

4. Trying to de-link selectivity and fecundity *without* HSPs leads to poor CVs (in line with the theoretical result that this would be impossible except for details of functional form). Sticking to POPs would mean sticking to this assumption.

5. By 2025, the retrospective CV for abundance in 2010 is approaching the "long-term retrospective CVs" of section 5 (whereas in 2018, the retrospective CV for 2010 is still substantially higher). For SBT CKMR, "the long term" therefore seems to be something like 10-15 years.

6. For the particular quantity considered here (N_{10+}), it does not matter much what the sampling (ie genotyping) ratio of J:A is, at least within the range 1:1 to 2:1. However, as noted earlier, it may be wiser to keep towards 1:1, in order to get more information on parents.

7. Conclusions and Suggestions

1. CKMR can be used effectively, and cheaply, as a long-term monitoring tool: long-term sampling levels can be reduced considerably compared to the 3000-ish-per-year SBT that were being genotyped.

    a) If abundance increases— at some rate $R\%$ per year, say— then CK sampling levels also need to increase to maintain CVs, but only at a rate $0.5R\%$ per year. Of course, the CK (and other) data itself will *tell* us that abundance is increasing in time to make this adjustment.

    b) Although this document considers medium- and long-term CVs from CKMR as the yardstick (with a number of caveats about the reliability of those CVs), it is obviously far beyond the scope to think what target CVs are needed for management.

2. It is time to reconsider the genetic technology used for SBT CKMR. Microsatellites are logistically fragile for long-term comparability, being dependent on key individuals' experience with SBT and on particular technology which cannot be guaranteed indefinitely into the future. They are also not going to become much cheaper. Of the two SNP-based alternatives, SNP-chips *might* be the cheapest right now (it's not completely clear), but DArTs are the most "future-proofed" (and reproducible across companies) with the least up-front cost, and their unit cost is likely to fall substantially with technological efficiencies. Changing the technology is going to become inevitable at some stage, and the longer it is left, the more duplicate expense (or foregone information) will be required later on.

    a) DArTs can reveal half-sibling pairs (HSPs) as well as POPs— though this does need to be checked on real samples. This is enormously powerful, both because it greatly increases the number of CKPs per sample, and also because HSPs (in different juvenile cohorts) allow the separate estimation of adult selectivity-at-age and fecundity-at-age. With POPs alone, it is necessary to *assume* constant daily-catchability across body-sizes in order to estimate adult abundance etc.

    b) Even at 2014 prices, DArTs are at least competitive with microsatellites because of the HSP factor, and will become more competitive over time.

3. In terms of a sampling and genotyping strategy for next few years, here is one suggestion:

    a) Continue to *collect* at current levels, 1500 juveniles and 1500 adults per year. Sampling and archiving is cheap, and keeps options open for the future.

    b) Start by re-genotyping 3000 juveniles from two years[16] in 2006-2010. This should yield around 25 cross-cohort HSPs— enough to check that DArTs are working OK for HSPs— and will allow streamlining of the SBT DArT process to cut subsequent costs. It will also allow a check on whether there are appreciable numbers of within-cohort HSPs in each juvenile sample.

    c) Aim by 2018 to genotype an annual *total* of 1500 juveniles plus adults, plus to work through the rest of the 2006:2010 back-catalogue of juveniles (another 5000 juveniles, once step (b) is done).

    d) The total number of animals to be genotyped over a 4-year-period would then be about 8000 + (2018-2011)*1500 = 18500 animals; note that there are already a lot of ungenotyped animals "accumulated" from 2011-2014 sampling. The timing of all the genotyping— apart from step (a)— doesn't really matter, and may to some extent depend on anticipated future costs.

    e) *If* the CV calculations in this paper are reliable, then this level of genotyping until 2018 would give a standard error on trend-in-log-abundance over the period 2006-2016— based entirely on "fishery-independent data" (ie no total catch or CPUE)— of about 0.09. In other words, if there is truly a 20% change in adult abundance over that period, then it ought to be statistically detectable at 95% significance.

    f) POPs provide direct information on adult fecundity-at-age. With HSPs too, it becomes possible to directly estimate adult *mortality*. Together, these may be helpful in reconstructing likely historical levels of reproductive output (from long before the CKMR study began).

4. Longer-term DArT genotyping at around the 1500-per-year level may be sufficient to maintain acceptable CVs. However, the target CV should be considered in the light of a more coherent framework for using the CKMR data.

[16] The two years should probably be an even number of years apart, to maximize HSPs from skip-spawners in this test phase.

a) The CKMR data actually has (at least) two purposes for CCSBT ESC: as an independent check on the rest of the assessment, not reliant on catch or CPUE data; and as a component of the OM, to be deliberately *integrated* with the other data. It is an open question as to what this should mean in terms of target CV, and what model should be considered when predicting what the CVs would be under different sampling strategies.

5. Continued collection of Indonesian length/sex/age data is essential for interpreting CKMR data.
6. To put future CKMR data into the OM, a sex-based and length-based framework is needed. That is the only coherent way to get full value-for-money out of these two high-quality datasets: CKMR, and the Indonesian age and length data.

References

Anderson E (2010). *Computational algorithms and user-friendly software for parentage-based tagging of Pacific salmonids [online]*. Tech. rep. Pacific Salmon Commission: Chinook Technical Committee (US Section).

Aranda G, Abascal FJ, Varela JL, and Medina A (2013). "Spawning behaviour and post-spawning migration patterns of Atlantic Bluefin Tuna (*Thunnus thynnus*) ascertained from Satellite Archival Tags". *PloS one* 8.10, e76445.

Bravington M, Grewe P, and Davies C (2014). *Fishery-independent estimate of spawning biomass of Southern Bluefin Tuna through identification of close-kin using genetic markers*. FRDC Report 2007/034. CSIRO, Australia.

CCSBT (2012). *CCSBT Report of the Seventeenth Meeting of the Extended Scientific Committee*. Tech. rep. Commission for the Conservation of Southern Bluefin Tuna.

– (2013). *CCSBT Report of the Eighteenth Meeting of the Extended Scientific Committee*. Tech. rep. Commission for the Conservation of Southern Bluefin Tuna.

Fernández ME, Goszczynski DE, Lirón JP, Villegas-Castagnasso EE, Carino MH, Ripoli MV, Rogberg-Muñoz A, Posik DM, Peral-García P, and Giovambattista G (2013). "Comparison of the effectiveness of microsatellites and SNP panels for genetic identification, traceability and assessment of parentage in an inbred Angus herd". *Genetics and molecular biology* 36.2, pp. 185–191.

Hauser L, Baird M, Hilborn R, Seeb LW, and Seeb JE (2011). "An empirical comparison of SNPs and microsatellites for parentage and kinship assignment in a wild sockeye salmon (*Oncorhynchus nerka*) population". *Molecular Ecology Resources* 11, pp. 150–161.

Hillary R, Preece A, Davies C, Bravington M, Eveson P, and Basson M (2012). *Initial exploration of options for inclusion of the close-kin data into the SBT operating model*. Scientific Committee report (OMMP workshop) CCSBT-ESC/1208/21. Commission for the Conservation of Southern Bluefin Tuna.

Seeb JE, Carvalho G, Hauser L, Naish K, Roberts S, and Seeb LW (2011). "Single-Nucleotide Polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms". *Molecular Ecology Resources* 11, pp. 1–8.

A. How POPs and HSPs can separate fecundity from selectivity

Consider a population in *quasi-equibrium*, ie with stable age structure over time, but not necessarily constant abundance (there can be exponential trend of increase or decrease). Assume a constant death rate δ over age

and time (δ replacing z here to be more Greek), and a constant "birth" rate β of adults in the youngest age class (which WLOG is here taken to be 0), so that the population rate-of-growth is $\rho \triangleq \beta - \delta$. It turns out that the stable age distribution is exponential with log-slope $-\beta$ (it is more usual to see $-\delta$, but that is only true for a population in *full* equilibrium, ie with stable abundance too). Also assume for convenience that fecundity is an exponential function of age a, proportional to $\exp(\phi a)$, and similarly that selectivity is proportional to[17] $\exp(\sigma a)$. The goal of the exercise is to estimate the four structural parameters$\{\beta, \delta, \phi, \sigma\}$.

In terms of data, assume we know the age composition of the adult catch, which will have slope $\sigma - \beta$, and that we know the age of all identified parents, and that all juveniles are caught soon after birth, and that we have sampled constant numbers of adults and juveniles per year for long enough to accumulate large number of POPs (and/or HSPs) available over a long period. The fundamental equation for further progress with POPs is the probability p_{aTt} that an adult captured at age a in year $T+t$ after the birth of a given juvenile in year T will be the juvenile's Parent. This is zero if $a < t$, or if not then it is given by the following formula, which is presented for (mathematical) convenience in continuous-time rather than discrete-time:

$$\begin{aligned} p_{aTt} &= \#\text{parents} \times \mathbb{P}[\text{parent age at birth was } a-t] \times \frac{\mathbb{P}[\text{parent survives } t \text{ years}]}{\#a\text{-years-olds alive in year } T+t} \\ &= 2 \times (\beta - \phi)^{-1} e^{(\phi-\beta)(a-t)} \times \frac{e^{-\delta t}}{N_0 e^{\rho(T+t)} \beta^{-1} e^{-\beta a}} \\ &= \frac{\beta}{\beta - \phi} \frac{2}{N_0} e^{-\rho T} e^{(\phi-\beta)a} e^{(\beta-\phi)t} e^{-(\rho+\delta)t} e^{\beta a} \\ &= \frac{\beta}{\beta - \phi} \frac{2}{N_0} e^{-\rho T} e^{-\phi t} e^{\phi a} \end{aligned} \tag{A.1}$$

This makes use of the (assumed) fact that all living adults of the same age (and sex) are equally likely to be caught.

It is evident from eqn (A.1) that there is a limited amount of information about structural parameters in the POPs:

- ρ (ie $\beta - \delta$) is identifiable from the slope with T (several years of data will be required to estimate it)
- otherwise, only ϕ appears in the slope terms (with a, t, and/or T).
- the term $\sigma - \beta$ is the slope of the adult "catch curve", but there is no data to separate the two

Another way to think about it, is to consider the properties of simple summary statistics in terms of the structural parameters:

1. Adult age in *catch*: slope $\sigma - \beta$, mean $(\beta - \sigma)^{-1}$
    a) σ because older adults are more likely to be caught
2. Mean parent age, if caught t years after Offspring birth: $t + (\phi + \beta - \sigma)^{-1}$
    a) ϕ because older parents make more offspring, so their offspring are "over-represented" in the sampled juveniles
3. Difference between (1) and (2) => ϕ
4. Rate-of-change of [eventual numbers of parents caught] with [year of offspring capture]: $-\rho$

[17] A few upper-bound constraints on ϕ and σ relative to the other parameters are required to avoid paradox, but do not affect the basic points.

a) because population is growing, so per-comparison chance of finding an individual parent is decreasing

So, although we can estimate ϕ, ρ, and $\sigma - \beta$, we can go no further with POPs alone.

From HSPs, the key quantity is the probability that a second juvenile J2, caught t years after the first juvenile J1 which was caught at year T, is indeed a half-sib of J1. It is clearest to calculate this separately for maternal and for paternal half-sibs, then add the two probabilities. To work out the probability, we need to integrate over the age-distribution of J1's mother at J1's birth. This age distribution is shifted towards older ages, because randomly-chosen juveniles is more likely to come from a more fecund (and therefore older) parent; it is exponential with slope $\phi - \beta$. Then the mother needs to survive t more years; finally, to find the probability that J2 will be one of her offspring, we need to compute her expected reproductive output in year t (by which time she will be older, and more fecund, than when she gave birth to J1) and divide it by the total reproductive output from all living females that year. This gives the following:

$$\begin{aligned}
&\mathbb{P}\left[\text{J2@}t\text{shares mother with J1@0}\right] \\
&= \frac{\int (\beta-\phi)^{-1} e^{(\phi-\beta)a} \times e^{-\delta t} \times e^{\phi(a+t)} da}{(N_0/2)\, e^{\rho(t+T)} \int \beta^{-1} e^{-\beta a'} \times e^{\phi a'} da'} \\
&= \frac{2}{N_0} \frac{\beta}{\beta-\phi} e^{-\rho T} e^{(\phi-\rho-\delta)t} \frac{\int e^{(2\phi-\beta)a} da}{\int e^{(\phi-\beta)a'} da'} \\
&= \frac{2}{N_0} \frac{\beta}{\beta-\phi} e^{-\rho T} e^{(\phi-\beta)t} \frac{\beta-\phi}{\beta-2\phi} \\
&= \frac{2}{N_0} \frac{\beta}{\beta-2\phi} e^{-\rho T} e^{(\phi-\beta)t}
\end{aligned} \tag{A.2}$$

5. From eqn (A.2), we can see that the mean gap between HSPs will be $(\beta-\phi)^{-1}$, so we can estimate $\beta \hat{-} \phi$.
6. Since we already know $\hat{\phi}$ from the POPs, we can estimate $\hat{\beta}$
7. Then we can add $\hat{\beta}$ to the slope of the adult age composition to estimate$\hat{\sigma}$
8. ... and we can subtract $\hat{\beta}$ from $\hat{\rho}$ to get $\hat{\delta}$

Note that eqn (A.2) also allows estimation of ρ, based on the rate-of-change of finding future HSPs per comparison as a function of T; and of N_0, based on the actual numbers of HSP found.

B. Quasi-equilibrium age distribution with senescence

[This was used to set up the AMA simulations for the long-term sampling targets— probably of interest only to the author.]

Suppose we are in quasi-equilibrium (stable age distribution, but possible exponential growth/decay in overall abundance), survivorship below plus group of $\exp(-\delta)$, plus-group survivorship $\exp(-\delta_+)$ at age A and above, and overall rate-of-increase $\exp(\rho)$. It's easily seen (as below) that the within-year slope of log-numbers-at-age below the plus-group is $-(\rho+\delta)$, but the proportion in the plus group is less obvious. Letting N_{at} be the

numbers at age a in year t, with $N_{A+,t}$ being the number in the plus-group, we have

$$N_{a+1,t+1} = e^{-\delta} N_{at} \text{ iff } a+1 < A$$
$$N_{a+1,t+1} = e^{\rho} N_{a+1,t}$$
$$\implies N_{a+1,t} = e^{-\rho-\delta} N_{at}$$
$$e^{\rho} N_{At} = N_{A,t+1} = e^{-\delta_+} N_{At} + e^{-\delta} N_{A-1,t}$$
$$\implies N_{At} = \frac{e^{-\delta}}{e^{\rho} - e^{-\delta_+}} N_{A-1,t}$$

C. Long-term theory of POP-based CK

For the cartoon version of CKMR, we know that m_J juvenile and m_A adult samples in a population of N adults give you $2m_J m_A/N$ expected POPs, and so to get 50 (which will give you a 15% CV) in the most economical way, you need $m_J = m_A$ and $m_J = 5\sqrt{N}$ so $m \triangleq m_J + m_A = 10\sqrt{N}$. But what about long-term monitoring? Intuitively, we expect that the annual sample requirements will be reduced once we have built up enough of a "catalogue" to get our first 50 POPs.

Start with the simplest case: an age-structured population but with knife-edge recruitment, constant mortality rate z thereafter, and constant fecundity-at-age. Although there will typically be several parameters other than N to be estimated, their contribution to overall CV of abundance will eventually become negligible as the total sample size increases.

We concentrate on estimating retrospectively the initial abundance R_T of the single *cohort* of adults that matures in year T. Once the structural parameters are effectively fixed, $\mathrm{CV}[R_T]$ will be determined entirely by the number of POPs P_T subsequently found for cohort T, via $\mathrm{CV}\left(\hat{R}_T\right) = 1/\sqrt{\bar{P}_T}$ where $\bar{P}_T \triangleq \mathbb{E}[P_T]$. So all we need to do for *that particular cohort,* is work out $\mathbb{E}[P_T]$ for different sampling strategies. Then we can combine cohorts— since the comparisons from different cohorts are independent— to find out the CV of things that are more directly useful, such as adult abundance in any given year. This approach transfers to more complicated scenarios where e.g. selectivity and fecundity are age-dependent.

Of course, in practice we may not know the age of each adult, just its length, so the tie-in to a specific cohort is inexact. However, if we are interested in an aggregate across cohorts, this doesn't matter; after all, the total number of POPs found is the same, so the "total information" is in some sense the same, and the split between cohorts is "just" an allocation issue.

Suppose we want the *lifetime accumulated POPs* for cohort T (i.e. adults maturing in year T), and assume the average size of a cohort is R, and neglect variability in total (adult) abundance caused by variable recruitment to individual cohorts. Ignore selectivity and fecundity changes within adulthood for now. Suppose also that juvenile age is always known. Let R be the initial size of a cohort when it matures, so that the total adult abundance is

$$N = \sum_{t'=0}^{\infty} R e^{-zt'} = R/\left(1 - e^{-z}\right)$$

Now consider a particular juvenile born in year $t+T$. The chance that it has a parent in cohort T is twice (♂ + ♀) the ratio of (# T's alive in $t+T$) to (# adults alive in $t+T$), ie $2e^{-zt}\left(1-e^{-z}\right)$. The chance of that parent

being caught, in year $t+T$ or subsequently (forgetting about the fiddly within-birth-year issue) is

$$\begin{aligned}\frac{k_A\left(1-e^{-z}\right)}{R}\times\sum_{s\geqslant t}e^{-z(s-t)}\\ =\frac{k_A}{R}\end{aligned}$$

ie the chance of finding that particular adult in year $t+T$ in the k_A samples from a total abundance $R/\left(1-e^{-z}\right)$, plus the chance of finding it in the next year given it might die, etc. This means the overall probability of matching that juvenile to a sampled adult from cohort T is

$$2e^{-zt}\left(1-e^{-z}\right)\times\frac{k_A}{R}$$

Now, there are k_J juveniles sampled every year, and to get the total number of POPs expected from cohort T, we sum this:

$$\begin{aligned}\bar{P_T}&=\frac{2k_Jk_A}{R}\left(1-e^{-z}\right)\sum_{t=0}^{\infty}e^{-zt}\\ &=\frac{2k_Jk_A}{R}\end{aligned}\tag{C.1}$$

Note that eqn (C.1) is a *retrospective* estimate; for the highest precision, we need to wait until there is negligible chance of finding more POPs from cohort T. Note also that, by Poissonness, $C\triangleq\mathrm{CV}\left[\hat{R}_T\right]=\sqrt{R/2k_Jk_A}$.

The total abundance across all cohorts in year T is

$$N_T\triangleq\sum_{s=0}^{\infty}e^{-zs}R_{T-s}$$

By the time we have "fully sampled" cohort T, we will also have fully sampled all the older cohorts already mature by T. Thus, the $\hat{R}_{T-s}$ are all independent with the same mean and variance. Thus

$$\begin{aligned}\mathbb{E}\left[\hat{N}_T\right]&=\bar{R}/\left(1-e^{-z}\right)\\ \mathbb{V}\left[\hat{N}_T\right]&=\mathbb{V}\left[\hat{R}_T\right]/\left(1-e^{-2z}\right)\\ \implies\mathrm{CV}\left[\hat{N}_T\right]&=\mathrm{CV}\left[\hat{R}_T\right]\frac{1-e^{-z}}{\sqrt{1-e^{-2z}}}\\ &\approx C\frac{z}{\sqrt{2z}}=C\sqrt{\frac{z}{2}}\end{aligned}\tag{C.2}$$

and remember that

$$\begin{aligned}C&=\sqrt{\bar{R}/2k_Jk_A}\\ &=\sqrt{\bar{N}\left(1-e^{-z}\right)/2k_Jk_A}\\ &\approx\sqrt{\frac{\bar{N}z}{2k_Jk_A}}\end{aligned}$$

Hence you get

$$\mathrm{CV}\left[\hat{N}_T\right] \approx \frac{z}{2}\sqrt{\frac{\bar{N}}{k_J k_A}}$$

C.1. Incorporating HSPs

HSPs yield an estimate of absolute adult abundance that is statistically independent of a POP-based estimate (given that the structural parameters are fixed, as they are by assumption in this long-term exercise). Since we do not know the age of the shared parent of an HSP, the information is localized to years, rather than to adult cohorts. The expected number of HSPs eventually found that have one member born in year Y is inverse to adult abundance in year Y, with known constant of proportionality that depends on the structural parameters. Thus we could form an HSP-based abundance estimate for Y that is statistically independent of the POP-based estimate above. To avoid double-counting HSPs (ie an HSP with members born in Y_1 and Y_2 does not contribute a full unit of information about abundance in *both* years), we halve the expected number of HSPs. An

$$\begin{aligned} v_P &\triangleq \mathbb{V}\left[\hat{N}_{\mathrm{POP}}\right] \\ v_H &\triangleq \mathbb{V}\left[\hat{N}_{\mathrm{HSP}}\right] \\ \hat{N} &= \frac{v_P^{-1}\hat{N}_{\mathrm{POP}} + v_H^{-1}\hat{N}_{\mathrm{HSP}}}{v_P^{-1} + v_H^{-1}} \\ \mathbb{V}\left[\hat{N}\right] &= \frac{1}{v_P^{-1} + v_H^{-1}} \end{aligned} \tag{C.3}$$

D. Finding HSPs with DArTs

Each SNP has two possible alleles[18]. An Unrelated Pair (UP) of animals will have an allele in common unless they are opposing homozygotes (ie one animal is AA and the other BB). If the MAF (allele frequency[19]) of one allele is f, then the probability of having an allele in common is $1-2f^2(1-f)^2$. Even for the most informative SNP possible, with MAF=0.5, this is still 7/8 or (56/64). Half-sibs are guaranteed to have an allele in common at any SNP that is co-inherited, which happens with probability 1/2, the non-co-inherited loci have the same allele-in-common probability as for an UP, so the overall proportion of shared-allele loci works out at 60/64 (MAF=0.5).

The simplest (not quite the most statistically efficient) way to distinguish UPs from HSPs is therefore to count the proportion of shared-allele SNPs in that pair; the aim is to decide whether the proportion comes from a distribution with mean 56/64 or mean 60/64 (actually these figures should use the real MAF, not 0.5), so clearly a large number of SNPs are required. The actual proportion in any UP will be roughly Binomially distributed[20],

[18] An "allele" here is just a DNA base: CAGT. Some rare SNPs have more than 2 possible alleles. We discard such SNPs.

[19] The literature uses the terms Minor Allele Frequency or Major Allele Frequency, according as f is below or above 0.5. Here I use MAF for either.

[20] Actually, the number of allele-in-common SNPs is not strictly Binomial, but rather a sum of independent Bernoullis, because the MAFs differ between SNPs. However, using an "average MAF" allows the much simpler Binomial calculation, which is quite adequate for crude design purposes.

The "independent SNPs" assumption for UPs eventually breaks down due to "linkage disequilibrium" if the number of SNPs is extremely high (say 10^5), so that the average physical spacing on the genome is small. This is irrelevant to our SBT DArT setting, where the number of SNPs is much lower. "Linkage disequilibrium" is not to be confused with the fact that SNP-inheritance in

and the the proportion in any HSP will be roughly an overdispersed Binomial, because SNPs are not inherited independently (genomes are inherited in contiguous chunks); the two distributions are bound to overlap slightly, and it is necessary to set a threshold proportion for classifying a pair as UP or HSP. The threshold can be chosen in advance based on the MAFs so as to ensure only a small number of "lucky UPs" fall above it and end up misclassified as HSPs.

SNPs with low MAFs are not very useful, but a preliminary scan through our SBT DArT results (over 37,000 SNPs in all) shows about 3000 SNPs with MAF$\geqslant$ 0.35, and 2000 with MAF$\geqslant$ 0.4. To check whether this will be enough to reliably distinguish HSPs, we make some conservative guesses:

- equivalent information of 2500 SNPs with average MAF=0.4 (we probably have more information than this).
- genotyping 1000 juveniles per year for 20 years (NB comparisons across more than 20 years can be ignored because few adults will live long enough).

This would lead to 2×10^8 pairwise comparisons, and maybe 500 true HSPs. We want to choose a threshold that will keep the expected number of "false-positive HSPs" (ie UPs that come above the threshold) down to say 1% of true HSPs, i.e. around 5 pairs. Inspection of the extreme tail of the Binomial distribution for UPs then implies that the threshold for "HSP classification" should be set around 2100 allele-sharing loci. The *expected* number of allele-sharing loci in a true HSP is 2356, well above the threshold. The proportion of true HSPs likely to be accidentally excluded by falling below the threshold cannot be reliably calculated in advance; it depends on the (unknown) chromosomal structure and (unknowable) crossover rate of the SBT genome. However, based *very roughly* on human genomic properties (20-ish chromosomes, average 1 crossover per chromosome per meiosis, two meioses per HSP), well under 1% of true HSPs would have to be sacrificed to the threshold.

After the event— i.e. by the time we have identified a respectable number of HSPs say 50— there will be enough information to estimate the overdispersion in proportion-of-allele-sharing loci among HSPs, and therefore to estimate the proportion of true HSPs sacrificed. If this is non-trivial, then it can be used to bias-adjust estimates of absolute abundance, which depend on the actual number of HSPs found.

As to the possibility of mis-identifying other relatives as HSPs: aside from full-sibs (very unlikely in different juvenile cohorts, and easily identified genetically) and half-sibs (the target), the next-closest relationship amongst juveniles is first-half-cousins (i.e. single shared grandparent). At MAF=0.5, FHCPs share an allele at, on average, 57/64 of loci, as opposed to 56/64 for UPs and 60/64 for HSPs. However, there are vastly fewer FHCP than UP, so the fact that 57 is slightly closer to 60 than 56 is, becomes irrelevant; if the threshold is tight enough to exclude stray UPs, then it will easily exclude the much rarer FHCPs from being mistakenly identified as HSPs.

D.1. Finding POPs with DArTs

With 2500 DArTs it is utterly trivial to find POPs. Because of the very large number of loci, some allowance must be made for genotyping errors (see below), by setting a threshold number of shared-allele loci such that any pair sharing at least that many are deemed a POP. (Without allowance for error or mutation, the threshold would be exactly equal to the number of loci.) Even allowing for an insanely slack 100 genotyping errors per pair of animals, the chance that a UP would share alleles at 2400 out of 2500 loci is below 10^{-40}.

close-kin will be correlated over *moderate* genomic distances (eg two SNPs in the same half of a chromosome), which is what generates overdispersion in HSPs. See genetics textbooks, and blame them for the terminology.

D.1.1. Finding POPs with a SNP-chip

With SNP-chips, the cost is proportional to the number of chips used, and each chip can accommodate around 100 (or sometimes 50) SNPs, so the total number of loci will be close to a multiple of 100 (or perhaps 50). Suppose we genotyped 2x100 SNPs at average MAF of 0.45 (the DArT results yielded nearly 1000 SNPs with MAF$\geqslant$ 0.45, though not all might work on SNP-chips) and again 2×10^8 adult-juvenile comparisons (1000 of each per year for 20 years) yielding several hundred POPs. To keep the number of false-positives below about 1% of true POPs, the threshold would need to be set at 196 loci— i.e. leaving room for 4 errors/mutations per pair. This is possibly OK but perhaps rather close for comfort; 250 loci with a margin-for-error of 7 loci might be wiser. Note that the conclusion is not particularly sensitive to sampling levels; if we "only" consider 2×10^7 comparisons, then this merely increases the margin size by one locus, because tail probabilities fall so fast.

D.2. Genotyping errors and mutations

Genotyping errors and mutations are of course possible, and with so many loci a few are inevitable. However, the literature I have seen suggests that SNP genotyping rates are generally below 1%, which will only cause a small shift in the expected allele-in-common proportions for HSPs and UPs, and is not going to suddenly cause massive overlap in the distributions. Unlike POPs, the process of finding HSPs does not use an exclusion principle— there is no requirement that *any* locus must have a shared allele, so the process is intrinsically somewhat robust to genotyping errors.

Given a known POP pair, any opposing homozygotes can only come from error/mutation. The SBT DArT data includes over 40 known POPs (based on usat data which is unlikely to include any false-positives pairs, and which for most pairs is completely unambiguous), ie 120000 comparisons between SNPs with MAFs over 0.35. This is an excellent test dataset for estimating the proportion of genotyping errors and/or mutation, and results should be available by CCSBT ESC 2014.

D.3. Maternal vs paternal HSPs

The expected number and pattern of HSPs depends on whether the shared parent is the Mother or the Father. To determine this, we can analyse the mitochondrial DNA (mtDNA) of all HSP fish. MtDNA is always inherited from the mother, so if the two HS have *different* mtDNA, they must have different mothers— ie their shared parent must be the father. On the other hand, if they have the *same* mtDNA, then they are likely to share a mother. However, there is a possibility that they actually share a father and just by chance happened to have different mothers with the same mtDNA. This probability $\mathbb{P}[\text{shared mother}|\text{HSP with same mtDNA}]$ can be calculated from mtDNA allele (AKA "haplotype") frequencies in the population. Once the probability is known, the HSP data can be analysed statistically using mixture-distribution ideas (i.e without requiring certainty about which parent was shared).

For this to work well, the mtDNA genome must have reasonably high diversity, so that we can reliably identify a good number of true *paternal* HSPs. Even if there is enough real diversity, we may not detect it unless the mtDNA genotyping process is sensitive enough[21]. In **Grewe1997**'s examination of SBT mtDNA, about 70% of fish all had particular one allele, i.e. an inconveniently low diversity. However, because the goal was traditional population genetics (and because of the analytical methods available at the time), the measurement

[21] Most genotyping of mtDNA uses only a small part of the mtDNA genome, and so will not differentiate between alleles (haplotypes) that differ only elsewhere in the mtDNA genome.

technique was deliberately chosen *not* to discriminate finely within alleles. Modern techniques can discriminate much more finely, and we have recently done this successfully for a species with a genuinely low-diversity mtDNA (and nuclear) genome; based on what we know of the SBT genome, we expect that its true mtDNA diversity is acceptably high. As long as the frequency of the commonest *measured* allele can be reduced to below, say, 0.4, there should not be any great loss of precision introduced by uncertainty about which parent is shared in some of the HSPs.

Genotyping the HSP mtDNA to establish maternal/paternal origin may cost up to $50 per fish, but since it is only necessary for fish in HSPs which will amount to a few hundred individuals, the additional cost is negligible.

E. A simplified framework for designing SBT CKMR

To do the calculations for this paper, it was necessary to implement a new "CK assessment model". The original CK Mini-Assessment (CKMA; CCSBT 2012 and 2013), which is fully length- as well as age-structured, is too complicated and slow for design purposes, has some structural constraints such as being unable to deal with changes in selectivity, and also gave baseline estimates with a much stronger trend[22] than the CK-OM combination model suggested. Since the recent trend has a big impact on future projections, it seemed wiser to start the projections with recent estimates from the CK-OM which (presumably) are better since they use other information. Some of the structural features of the Mini-Assessment may also be inadequate for long-term work, e.g. constant selectivity over time. Unfortunately, the CK-OM itself is not suitable for CK projections either; the most obvious problem is that it lumps both sexes together, despite plentiful sex data, clear evidence of dimorphic growth, and every reason to expect different selectivity and fecundity patterns by sex!

Instead, for this paper I developed a new Approximate Mini-Assessment (AMA) for adult population dynamics only, using just CK data and adult age data. It is purely age-and-sex-based model, with constant mortality z over time and over ages 8-24, a plus-group at[23] 25 with its own z_+, fixed sex-ratio over time, and sex-specific selectivity and fecundity curves.

To seed the model, I used the selectivity-at-length and fecundity-at-length estimates from the CKMA, converted from length to age. However, I also used the fitted age-composition data from the CK-OM, adjusted slightly in a least-squaresy way to match the detailed assumptions of the AMA (e.g. a constant z over ages 8-24, estimated from the CK-OM results). These choices are somewhat internally inconsistent; the CK-OM z is generally higher than the CKMA z, and would imply different selectivity and fecundity curves. I also adjusted the overall abundance to match the existing total of 45 POPs by 2010, since the untweaked AMA suggested rather more POPs (63) whereas, for purposes of assessing future CVs, we *know* we are starting with 45 POPs.

There is a fast and straightforward way (for statisticians...) to evaluate the information content of purely CK data; each pairwise comparison makes an independent contribution to the overall likelihood, and is in effect a Poisson-distributed RV with very low expected value[24], and its expected Fisher information ($d^2\text{loglik}/d\text{params}^2$) is easily enough computed. We need only simulate exact population dynamics, evaluate the probability of each comparison yielding a POP (or an HSP, as appropriate), count the number of comparisons of each type under a given sampling design, add up all the expected Fisher information matrices, and invert it to get the covariance matrix of all the parameters.

[22] The CK mini-assessment had considerable uncertainty about the *trend* in abundance over 2003-2010, though less so about the mean.

[23] To match the original choice in the CKMA, so that no extrapolation of previous results was required. With hindsight, 30 would have been a better choice in the original CKMA.

[24] Actually, each comparison is of course Bernoulli (0/1) not Poisson, but since the expected values are on the order of 10^{-8}, the distinction is irrelevant!

The more difficult problem is what to do about the Length- and Age-composition (L&A) data from Indonesia, which is *essential* to estimating abundance and everything else with CKMR data. In the CKMA, the L&A data were very informative, and allowing for uncertainty in them added remarkably little to the CV of abundance. Transition to a purely age-based setting is one tricky aspect, since only about 1/3 of adults are directly aged (although all identified parents are aged). In the AMA, age is assumed to be accurately known for all sampled adults (not just parents), so that every POP-seeking comparison is conditioned on adult age; although this is not realistic, all "real" adults are lengthed and sexed, and it is length rather than age that is likely to be the primary driver of selectivity and fecundity, so the "accurate age" assumption is a reasonable surrogate in a non-length-based formulation like AMA.

However, it seems unwise to treat age compositions as truly exact, in that it might interact subtly with detailed assumptions, e.g. about the shape of selectivity curves as opposed to the general trend with age, so as to provide "spurious information" about certain parameters. The compromise I used was to assume that the L&A data is accurate enough to reconstruct all the numbers-at-age-and-time (relative to some absolute abundance) *given* any set of parameters for mortality and selectivity. Thus the L&A data are assumed to provide *zero* information in themselves about mortality or selectivity. However, once those parameters (and the overall abundance) are set, the L&A data entirely determine the numbers-at-age-and-time, from which the probabilities of finding CKPs are computed. Operationally, this is done by using the true selectivity (in the simulation) and the true population-numbers-at-age to fix the catch-proportion-at-age in Year 1, and the catch-proportion-at-age-8 (the youngest adults) in each subsequent year. Then, given some trial selectivity parameters during estimation (and a trial initial abundance), the catch-proportion-at-age in Year 1 is used to determine the corresponding initial numbers-at-age. These population numbers-at-age are projected to year 2 using mortality, and the incoming 8yo are set based on the catch-proporation-at-age-8 in year 2, and the selectivities. This process is projected forwards to fill in the entire numbers-at-age matrix. This *notionally* is supposed to correspond to the L&A data providing enough information to "back-converge" all cohorts to their initial relative abundances, but not to overfit to details of changing age distributions in the catch from year-to-year that result from the interaction of selectivity and mortality.

I have been unable to think of any better yet practical way to handle the information associated with L&A, but it has to be admitted that this approach has flaws: it is too optimistic in some ways and too pessimistic in others. It is optimistic in the sense that, for making inferences near the end of the projections, the relative abundance of recently-recruited cohorts (ie cohorts currently not much older than 8) will *not* be known accurately; retrospective accuracy is only a reasonable approximation for cohorts that have been sampled as adults for enough years to accumulate reasonable sample sizes. [Having said that, in a "real" assessment that used pre-adult data as well, there would be prior information on the relative strength of each cohort as it nears adulthood.] By the same token, though,the AMA makes no assumption at all about trends in recruitment[25], which gives perhaps an unreasonable amount of freedom. Also, the AMA treatment of L&A also ignores detailed information about changing length-distributions which (given reasonable assumptions about length-at-age) may convey some real information about selectivity. But, on yet another hand, it is also not reasonable to assume over the long term that selectivity is constant, so in practice much of the information in detailed length data would be used to patch up short-term changes in selectivity. It is, in a word, complicated.

From experimentation, the L&A treatment leads to increased CVs relative to the CKMA in the short term (eg using only the existing data up to 2010). Then again, the CVs in the CKMA were remarkably close to the theoretical minimum imposed by the number of POPs found; perhaps some hidden structural assumption (e.g.

[25] Whereas the CKMA assumed no *expected* trend 2003-2010, though random fluctuations from year-to-year could still induce a *realized* trend.

about recruitment trends, or constant selectivity over time) was artificially lowering the CVs in the CKMA, and perhaps the CKMA CVs should not be over-trusted as a benchmark. Anyway, in the long term (20+ years), CVs from the AMA actually become implausibly low, this time for a clear reason: given the parameters for selectivity, fecundity, and mortality, only one single abundance parameter is required, and thereafter the L&A data is assumed to be capable of accurately setting the *entire* time-series of numbers-at-age across *any* period. That is asking too much of L&A data.

Given the way the AMA is built, its CVs ought to appropriately reflect the information content of the CKMR data itself, but not necessarily the additional— and very important— information content from the L&A data. Overall, I suspect (and within this study I can do no more than suspect) that the predicted CVs for different sampling strategies over the same period are reasonable at least in relative terms, and are not unreasonable in absolute terms over the the medium-term: say, up to 10 years hence. Beyond that, a more reliable way to make inferences about long-term CVs may be to base them "from first principles" on likely numbers of CKPs, as in Section 5. An even more reliable way would be to genotype some more fish over the next few years, and then reconsider.